

# Link-space formalism for network analysis

David M.D. Smith<sup>1,2,\*</sup>, Chiu Fan Lee<sup>1</sup>, Jukka-Pekka Onnela<sup>1,2,3</sup>, and Neil F. Johnson<sup>4</sup>

<sup>1</sup>*Physics Department, Clarendon Laboratory, Oxford University, Oxford OX1 3PU, U.K.*

<sup>2</sup>*Saïd Business School, Oxford University, Oxford OX1 1HP, U.K.*

<sup>3</sup>*Laboratory of Computational Engineering, Helsinki University of Technology, Finland and*

<sup>4</sup>*University of Miami, Coral Gables, Florida FL 33124, USA*

(Dated: February 4, 2008)

We introduce the link-space formalism for analyzing network models with degree-degree correlations. The formalism is based on a statistical description of the *fraction of links*  $l_{i,j}$  connecting nodes of degrees  $i$  and  $j$ . To demonstrate its use, we apply the framework to some pedagogical network models, namely, random-attachment, Barabási-Albert preferential attachment and the classical Erdős and Rényi random graph. For these three models the link-space matrix can be solved analytically. We apply the formalism to a simple one-parameter growing network model whose numerical solution exemplifies the effect of degree-degree correlations for the resulting degree distribution. We also employ the formalism to derive the degree distributions of two very simple network decay models, more specifically, that of random link deletion and random node deletion. The formalism allows detailed analysis of the correlations within networks and we also employ it to derive the form of a perfectly non-assortative network for arbitrary degree distribution.

PACS numbers: 89.75.Fb, 87.75.Hc, 05.40.-a

Networks – in particular large networks with many nodes and links – are attracting widespread attention. The classic reviews [1, 2, 3] with their primary focus on structural properties have been followed up by more recent ones addressing the role of dynamics, such as spreading and synchronization processes on networks, as well as the role of weights and mesoscopic structures, i.e. cliques and communities, within networks [4, 5]. Although several different measures for characterizing networks have been presented, for example in a recent survey [6], the simple concept of vertex degree remains unrivalled in its ability to capture fundamental network properties. When comparing the degrees of connected vertices, however, one often finds that they are correlated, a quality that gives rise to a rich set of phenomena. Degree correlations constitute a central role in network characterization and modelling but, in addition to being important in their own right, also have substantial consequences for dynamical processes unfolding on networks. Given the increasing current interest in network dynamics, understanding structural correlations remains important and timely.

In this paper we provide a detailed mathematical formalism for modelling degree-degree correlations within stochastically evolving, non-equilibrium networks. It is built around a statistical description of inter-node linkages as opposed to single-node degrees. While correlations have been characterized in empirical and model networks, most works devoted to analytical calculations of correlations in models, as pointed out in [7], have been performed only for particular cases. We start by providing a brief overview of degree correlations for network structure and dynamics in Section I and discuss how the

work presented relates to existing studies. The *link-space formalism*, which lies at the core of this paper, is introduced in Section II. The formalism comprises a master equation description of the evolution of a specific matrix construction, termed the *link-space matrix*, which characterizes degree-degree correlations within a network. The formalism can be implemented in a variety of ways. To demonstrate its use, we apply it to two well-known, non-equilibrium examples, namely random-attachment and Barabási-Albert (BA) preferential-attachment networks [8, 9] in Section II, and solve the so-called link-space matrix analytically for the steady state of these systems. In Section II, we also apply it to the classical equilibrium random graph of Erdős and Rényi [10] (ER) which, interestingly, requires a full, time-dependent solution of the link-space master equations. The cumulative link-space introduced in Section II aids comparison between simulated and analytically-derived link-space matrices and could, in principle, be applied to empirical networks to better ascertain their degree correlations. Analytic derivation of link-space matrices allows detailed analysis of the degree-degree correlations present within networks as is demonstrated in Section III. Within Section III, we also show how, for an arbitrary degree distribution, a link-space matrix can be derived which has no correlations present, i.e. the link-space is representative of a perfectly non-assortative network with that degree distribution. We then consider the counter-intuitive prospect of finding steady states of decaying networks in Section IV. In Section V, we introduce a simple one-parameter network growth algorithm, which is able to produce networks whose degree distributions exhibit a wide range of power-law exponents via a redirection process similar to the model by Krapivsky and Redner [11]. The model is interesting in its own right in the sense that it makes use of only local information about node degrees. Here

---

\*Electronic address: d.smith3@physics.ox.ac.uk

the link-space formalism allows us to identify the transition point at which higher exponent degree distributions switch to lower exponent distributions with respect to the BA model. We conclude in Section VI. To maintain readability, we postpone the more detailed mathematical derivations to Appendices A, B and C, and refer to them where appropriate.

## I. OVERVIEW OF DEGREE CORRELATIONS

A discussion of correlations might naturally start with the phenomenon of *clustering*. The highly influential paper of Barabási and Albert and many subsequent papers focussed on the modelling and analysis of models which replicate power-law degree distributions observed in many real-world systems [3, 8, 12]. However, the BA preferential attachment mechanism (discussed in Section II B) lacks certain key features observed in many of those systems, one such feature being clustering. Clustering is an important local statistic in many networks and reflects the connectedness of neighbours of a node. In a social context, if person  $Y$  is friends with both  $X$  and  $Z$ , one might expect some kind of link between  $X$  and  $Z$ . The measurement of the connectedness of neighbours of nodes is the clustering coefficient. This is often averaged over all nodes and a single value is given as the average clustering coefficient of the network [2]. Some node labelled  $\eta$  in an arbitrary undirected network has  $k_\eta$  neighbours, between which there could be a possible  $k_\eta(k_\eta - 1)/2$  links. If  $y$  of these are actually present, the clustering of node  $\eta$  is given as  $C = 2y/(k_\eta(k_\eta - 1))$ . The clustering nature of a network can also be expressed as the average over all nodes of degree  $k$  giving a clustering distribution (or *spectrum*),  $\overline{C}_k$ . As noted by Klemm and Eguíluz, in a BA network, the average clustering of a given node is independent of its degree [13], in contrast to the findings of Fronczak et al. [14], and tends to zero in the large-size (thermodynamic) limit. A network model which does exhibit high clustering is that of Watts and Strogatz in which a random rewiring process is carried out on an initially regular lattice [15]. The networks generated by such a process feature short average shortest path lengths between node pairs (the *small-world* effect). However, these networks do not exhibit a power-law degree distribution. Subsequently, there have been many efforts to build models which can encompass both of these features such as the non-equilibrium growing network model proposed by Klemm and Eguíluz which features deactivation of a nodes availability to be connected to [13]. Other models such as that of Holme and Kim [16] and that introduced by Toivonen et al. [17] modified the original BA preferential attachment mechanism, allowing further links between the new node and neighbours of the preferentially selected node. The social network model of Toivonen et al. produced communities with dense internal connections. Szabó et al. formulated a scaling assumption and a mean-field theory of cluster-

ing in growing scale-free networks and applied it to the Holme and Kim mechanism [18]. As discussed by Boguñá and Pastor-Satorras, clustering in networks is closely related to degree correlations [19]. Infact, based on the work of Szabó et al., Barrat and Pastor-Satorras introduced a framework for computing the rate equation for two vertex correlations in the continuous degree and continuous time approximation [7]. We shall now describe these degree correlations.

Vertex degree correlations are measures of the statistical dependence of the degrees of neighbouring vertices in a network. In general,  $n$ -vertex degree correlations, or  $n$  point correlations, can be fully characterized by the conditional probability distribution  $P(k_1, k_2, \dots, k_n | k = n)$  that a vertex of degree  $k = n$  is connected to a set of  $n$  vertices with degrees  $k_1, k_2, \dots, k_n$ . Two and three-point correlations are of particular interest in complex networks as they can be related to network assortativity and clustering, respectively. More specifically, two vertex degree correlations (two point correlations) can be expressed as conditional probability  $P(k' | k)$  that a vertex of degree  $k$  is connected to a vertex of degree  $k'$ . Similarly, three vertex degree correlations (three point correlations) can be fully characterized by the conditional probability distribution  $P(k', k'' | k)$  that a vertex of degree  $k$  is connected to both a vertex of degree  $k'$  and a vertex of degree  $k''$ . This implies that the degrees of neighbouring nodes are not statistically independent. Reliable estimation of  $P(k' | k)$  and  $P(k', k'' | k)$  requires a large amount of data and, in practice, one often resorts to related measures. Instead of  $P(k' | k)$ , the average degree of nearest neighbours,  $\langle k_{nn} \rangle_k$ , of nodes with degree  $k$  is often measured. This can be formally related to  $P(k' | k)$  [19, 20]. If  $\langle k_{nn} \rangle_k$  increases with  $k$ , high degree vertices tend to connect to high degree vertices. A network with this property is described as *assortative* or displaying positive degree correlations. If  $\langle k_{nn} \rangle_k$  decreases with  $k$ , high degree vertices tend to connect to low degree vertices (*disassortative* or negatively correlated) [19, 21]. An alternative to  $\langle k_{nn} \rangle_k$  is to use a normalised Pearson's correlation coefficient of adjacent vertex degrees providing a single number measure of assortativity as suggested by Newman to further classify networks [3, 21]. These approaches are discussed in more detail in Section III. To characterize three point correlations, instead of using  $P(k', k'' | k)$ , one can employ the clustering spectrum  $\overline{C}_k$ , the average clustering coefficient of nodes with degree  $k$ , which can be related to  $P(k', k'' | k)$  [19]. In many real-world networks such as the Internet [20], the clustering spectrum is a decreasing function of degree and while this is sometimes interpreted as a signature of hierarchical structure in a network, Soffer and Vázquez suggested that this is a consequence of degree-degree (two point) correlations that enter the definition of the standard clustering coefficient [22]. The authors introduced a different definition for the clustering coefficient that does not have the degree-correlation 'bias', i.e. a three point correlation measure that filters out two point correlations. Following the suggestion of

Maslov *et al.* [23] that these phenomena might arise from topological constraints rather than evolutionary mechanisms, Park and Newman demonstrated that disassortative degree correlations observed in the Internet could be explained via the restriction of there being no double edges between nodes [24]. In contrast, social networks have been found to be assortative [17]. Similarly, Catazaro *et al.* observed that the network of scientific collaborations was assortative and presented a model to reproduce this feature [25].

Functional processes occurring on networks are influenced by degree correlations, highlighting the importance of their role in complex networks. Eguíluz and Klemm considered highly clustered scale-free networks and showed that correlations play an important role in epidemic spreading [26]. The time average of the fraction of infected individuals in the steady-state undergoes a phase transition at a finite critical infection probability. They related this critical threshold to the transmission probability and the mean degree of nearest neighbours of all nodes in the system,  $\langle k_{nn} \rangle$ , the conjectured criterion for epidemic spreading being related to the product of the two. The value  $\langle k_{nn} \rangle$  scales with the system size in their highly clustered scale-free network more slowly than in the random scale-free which is a byproduct of the disassortativity of the system. Consequently, whereas in random scale-free networks in which viruses with extremely low spreading probabilities can prevail, the absence of connections between highly connected nodes in highly clustered scale-free networks protects the system against epidemics [26]. Interestingly, Boguñá *et al.* assert that any scale-free network of appropriate exponent will have diverging  $\langle k_{nn} \rangle$  in the thermodynamic limit, resulting in no threshold properties for epidemic spreading regardless of the correlations within the network [27] in contrast to the earlier suggestion of Boguñá and Pastor-Satorras [28]. Brede and Sinha induced correlations to Erdős-Rényi and scale-free networks by rewiring them appropriately to examine their dynamic stability. They mapped the adjacency matrices into Jacobian matrices and examined the largest eigenvectors, reflecting the decay rates of perturbations about an equilibrium state. They found that positive correlations reduced their dynamical stability [29]. Similarly, Bernardo *et al.* induced negative degree correlations in scale-free networks whose links couple non-linear oscillators. Through analysis of the eigenratio of the Laplacians of such networks, they found that network synchronisability improved as the network was made more disassortative [30]. Based on this result, they conjectured that negative degree correlations may emerge spontaneously as the networked system attempts to become more stable [31]. The same authors found similar results to hold also for weighted networks [30]. Maslov and Sneppen found that in “interaction and regulatory networks, links between highly connected proteins were systematically suppressed, whereas those between highly connected and low-connected pairs of proteins were favoured”, a topological organisation that in-

creases the overall robustness of the network to perturbations through “localising deleterious perturbations” [32]. This is consistent with the findings of Berg *et al.* whose model of the evolution of protein interaction networks exhibited disassortativity consistent with their empirical findings [33].

Krapivsky *et al.* used a master equation method, in which rate equations for the densities of nodes of a given degree are employed, to investigate the steady state of the BA preferential attachment mechanism [34]. As illustrated in Section II, this is a general method which can be applied to various growing network models. To extend the method to encompass two-point correlations, Krapivsky and Redner applied the approach to the number  $N_{k,l}$  of nodes with total degree  $k$  connected to ancestor nodes of total degree  $l$  in a *directed* network [11]. They solved analytically the master equations for the steady-state of a specific directed network growth model, namely the Growing with Redirection algorithm [11], similar to the mixture model introduced in Section V. Boguñá and Pastor-Satorras considered a link orientated description of two point correlations. For *undirected* graphs, they introduced a symmetric matrix whose elements  $E_{k,l}$  represent the number of links connecting nodes degree  $k$  to nodes degree  $l$  [19]. Boguñá and Pastor-Satorras related this matrix to the joint degree distribution  $P(k' | k)$  but noted that finite size effects make empirical evaluation of the matrix difficult, suggesting the use of the spectrum  $\langle k_{nn} \rangle_k$  as a more suitable observable. Here, we introduce a similar matrix construction which can be applied to both the *undirected* and *directed* scenarios. We call this the *link-space* matrix. We show that it is possible to construct master equations to model the evolution of this matrix which can be applied to a wide variety of network evolution algorithms retaining important degree correlations which can be critical to the network’s development. This framework is termed the *link-space formalism* and is introduced in Section II. In certain cases, these master equations can be solved analytically providing a full time-dependent solution or a steady-state solution of the link-space matrix of the network. From these analytically derived link-space matrices both the degree and joint degree distributions can be obtained allowing accurate analysis of degree correlations. The formalism also allows the derivation of the form of networks of predetermined correlation properties such as a perfectly non-assortative network (Section III A) and the derivation of the steady-state solutions to network decay algorithms (Section IV). The formalism can also be employed in its iterative guise to provide approximations to the steady-state when an analytic solution is not possible such as for the mixture model of Section V.

One and two-point correlations have natural physical counterparts within the network, specifically the nodes and links. The one-point correlations  $P(k)$  are simply related to the fraction of nodes in the network with degree  $k$ . As described in detail within this paper, two-point

correlations  $P(k' | k)$  are related to the fraction of links within a network connecting nodes of degree  $k$  with nodes of degree  $k'$ , hence the term *link-space*. One can extend the analysis to three-point correlations,  $P(k', k'' | k)$ . This would be related to the number (fraction) of pairs of links within a network sharing a common node of degree  $k$ , the remaining link ends being connected to nodes of degrees  $k'$  and  $k''$  and all possible ‘open triangles’ within the network would have to be considered. Clearly, the process can be extended to arbitrary  $n$  point correlations although the physical interpretation of the appropriate measurable quantities will become increasingly obscure.

## II. NODE-SPACE AND LINK-SPACE

We now introduce the link-space formalism. Consider a simple, growing, non-equilibrium network in which one node is added to a network at each timestep and this node is connected to the existing network with exactly  $m$  undirected links. The process is governed by an attachment probability *kernel*  $\Theta_j$ , defined as the probability that a specific, newly-introduced link attaches to any node of degree  $j$  within the existing network. At some time  $t$  there exist  $X_i(t)$  nodes of degree  $i$  and we wish to compute the expected number of nodes with degree  $i$  at time  $t + 1$ . The *node-space master equations* can be expressed in terms of the attachment kernels and are written

$$\begin{aligned}\langle X_i(t+1) \rangle &= X_i(t) + m\Theta_{i-1}(t) - m\Theta_i(t) \quad i > m, \\ \langle X_m(t+1) \rangle &= X_m(t) + 1 - m\Theta_m(t),\end{aligned}\quad (1)$$

since a new node of degree  $m$  is added to the existing network at each timestep and there are no nodes with degree less than  $m$ .

So far we have said nothing about the attachment mechanism, and have made the easily generalizable restriction that only one node is being added per timestep with undirected links. We now follow a similar analysis, but retain the node-node linkage correlations that are inherent in many real-world systems [11, 21, 35]. Consider any link in a general network – we can describe it by the degrees of the two nodes that it connects. Hence we can construct a matrix  $\mathbf{L}(t)$  such that the element  $L_{i,j}(t)$  is equal to the number of links from nodes of degree  $i$  to nodes of degree  $j$  for  $i \neq j$  at some time  $t$ . To ease the mathematical analysis below, the diagonal element  $L_{i,i}(t)$  is defined to be *twice* the number of links between nodes of degree  $i$  for the undirected graph, a mathematical convenience also observed by Boguñá and Pastor-Satorras [19]. For undirected networks  $\mathbf{L}(t)$  is symmetric and  $\sum_{i,j} L_{i,j}(t) = 2M(t)$ , twice the total number of links  $M(t)$  in the network which is simply  $mt$  when introducing  $m$  links per timestep. The matrix,  $\mathbf{L}$ , represents a surface describing degree-degree correlations in the network (see Figs. 1, 2 and 3) and is called the *link-space* matrix.

Consider one of the newly introduced links, one end of which is attached to the new node. The probability

of selecting any node of degree  $i - 1$  within the existing network for the other end to attach to is given by the attachment probability  $\Theta_{i-1}(t)$ . Suppose an  $i - 1$  node is selected. The fraction of nodes of degree  $i - 1$  that are connected to nodes of degree  $j$  is

$$\frac{L_{i-1,j}(t)}{(i-1)X_{i-1}(t)}.$$

The expected increase in links from nodes of degree  $i$  to nodes of degree  $j$ , through the attachment of the new node to a node of degree  $i - 1$ , is given by

$$\frac{\Theta_{i-1}(t)L_{i-1,j}(t)}{X_{i-1}(t)}.$$

Since each link has two ends, the value  $L_{i,j}$  can increase by a connection to an  $(i - 1)$ -degree node which is in turn connected to a  $j$ -degree node, or by connection to an  $(j - 1)$ -degree node which is in turn connected to an  $i$ -degree node. We write master equations governing the evolution of the link-space matrix as the evolution of the expected number of links from  $i$  to  $j$  degree nodes, i.e. the number of  $i \leftrightarrow j$  links. This is the *link-space formalism* and is written for the (generalizable) case of adding one new node with  $m$  undirected links to the existing network as

$$\begin{aligned}\langle L_{i,j}(t+1) \rangle &= L_{i,j}(t) + \frac{m\Theta_{i-1}(t)L_{i-1,j}(t)}{X_{i-1}(t)} \\ &\quad + \frac{m\Theta_{j-1}(t)L_{i,j-1}(t)}{X_{j-1}(t)} - \frac{m\Theta_i(t)L_{i,j}(t)}{X_i(t)} \\ &\quad - \frac{m\Theta_j(t)L_{i,j}(t)}{X_j(t)}, \quad i, j > m, \\ \langle L_{m,j}(t+1) \rangle &= L_{m,j}(t) + m\Theta_{j-1}(t) + \frac{m\Theta_{j-1}(t)L_{m,j-1}(t)}{X_{j-1}(t)} \\ &\quad - \frac{m\Theta_m(t)L_{m,j}(t)}{X_m(t)} \\ &\quad - \frac{m\Theta_j(t)L_{m,j}(t)}{X_j(t)}, \quad j > m.\end{aligned}\quad (2)$$

There are a variety of ways in which both the node-space and link-space master equations can be approached. For example, a full, time-dependent solution could be investigated as in Section II C or, using appropriate initial conditions, the equations can be iterated over the required timescale as in Section V. We can also investigate the possibility of a steady state of the algorithm under scrutiny. To do so, we assume that there exists a steady state in which the degree distribution remains static and investigate a solution [47]. Under this assumption, the fraction of nodes  $c_i(t) = X_i(t)/N(t)$  which have a given degree remains constant such that  $\langle X_i(t+1) \rangle - X_i(t) \approx dX_i/dt = c_i$  when one new node is added per timestep ( $N(t) = t$ ). It is also assumed that in the steady state, the attachment kernels are static too. We drop the notation ‘(t)’ to indicate the steady-state

and can rewrite Eq. (1) as

$$\begin{aligned} c_i &= m\Theta_{i-1} - m\Theta_i, \quad i > m, \\ c_m &= 1 - m\Theta_m. \end{aligned} \quad (3)$$

The fraction of links between nodes of degree  $i$  and nodes of degree  $j$  can be expressed as the normalized link-space matrix,  $l_{i,j}(t) = L_{i,j}(t)/M(t)$ , which sums to 2 in the undirected case. In the steady state we can assume that these values are static and can rewrite the link-space master equation (Eq. (2)) as

$$\begin{aligned} l_{i,j} &= \frac{\frac{\Theta_{i-1}}{c_{i-1}}l_{i-1,j} + \frac{\Theta_{j-1}}{c_{j-1}}l_{i,j-1}}{\frac{1}{m} + \frac{\Theta_i}{c_i} + \frac{\Theta_j}{c_j}}, \quad i, j > m, \\ l_{m,j} &= \frac{\frac{\Theta_{j-1}}{c_{j-1}}l_{m,j-1} + \frac{\Theta_{j-1}}{m}}{\frac{1}{m} + \frac{\Theta_m}{c_m} + \frac{\Theta_j}{c_j}}, \quad j > m. \end{aligned} \quad (4)$$

The notation ‘(t)’ has again been dropped to indicate the steady-state and here the generalizable situation of adding one new node per timestep is considered.

To apply the link-space formalism one starts by specifying the model dependent attachment kernel  $\Theta_i$ . To investigate a time-dependent solution, the attachment kernel is substituted into Eqs. (1) and (2). To investigate a steady-state solution, the kernel is substituted into the Eqs. (3) and (4) to yield recurrence relations for  $c_i$  and  $l_{i,j}$  respectively which can be solved analytically in some cases. The number of  $i$ -degree nodes is given by  $X_i(t) = i^{-1} \sum_k L_{i,k}(t)$  which allows us to retrieve the degree distribution from the normalized link-space matrix:

$$c_i(t) = \frac{M(t) \sum_{k=1}^{\infty} l_{i,k}(t)}{iN(t)}. \quad (5)$$

Degree distributions of empirically observed or simulated networks typically become dominated by noise at large degrees reflecting the small probabilities associated with these values occurring. In the link-space, the situation is exacerbated and the high  $i, j$  limit reflects connections between these high degree nodes. This is, of course, rarer than the existence of nodes of either degree. Following the conventional approach which is applied to degree distributions [2], we can use a cumulative representation of the link-space to address this issue. The use of a cumulative binning technique averages over stochasticity in the system. With degree distributions, regression techniques (curve fitting) applied to the cumulative distribution is used to obtain a more accurate description of the actual degree distribution than would be obtained from fitting to the empirical distribution itself [2]. The process can be similarly applied in the link-space. Surface fitting could be applied to the cumulative link-space obtained from an empirical network. From the cumulative fitted surface, a more accurate representation of the actual link-space could be obtained for the empirical network, from which a better representation of the network’s correlations could be obtained. This process also allows for

comparison between simulated and analytically-derived link-space matrices. We define the cumulative link-space matrix,  $cuml_{i,j}$  to be

$$cuml_{i,j} = \sum_{x=i}^{\infty} \sum_{y=j}^{\infty} l_{x,y}. \quad (6)$$

Note that we have not lost generality in that given a cumulative linkspace matrix, the actual link-space can be derived using variations on the following:

$$l_{i,j} = cuml_{i,j} - cuml_{i+1,j} - cuml_{i,j+1} + cuml_{i+1,j+1} \quad (7)$$

The computation involved when evaluating  $cuml_{i,j}$  can be cut down considerably by first evaluating  $cuml_{1,1}$ , which in the undirected graph is equal to 2. The leftmost column (or top row) for  $i > 1$  can then be evaluated as

$$cuml_{i,1} = cuml_{i-1,1} - \sum_{x=1}^{\infty} l_{i-1,x}. \quad (8)$$

Note that the second term on the right hand side is a row sum of the normalized link-space matrix and, hence, quickly evaluated. Indeed this row sum can be related to the degree distribution  $c_{i-1}$  from Eq. (5). Subsequent elements can be evaluated using the following simple identity,

$$\begin{aligned} cuml_{i,j} &= cuml_{i-1,j} + cuml_{i,j-1} \\ &\quad - cuml_{i-1,j-1} + l_{i-1,j-1}. \end{aligned} \quad (9)$$

We will now demonstrate the use of the link-space formalism to study a random attachment model and the Barabási-Albert (BA) model using steady-state solutions and the classical Erdős and Rényi random graph using a time-dependent solution.

#### A. Random attachment model (steady-state solution).

Consider first a random-attachment model in which at each timestep a new node is added to the network and connected to an existing node with uniform probability without any preference (‘Model A’ in [8, 9]) with one undirected link such that  $m = 1$  and  $M(t) \approx N(t) = t$ . We assume a steady-state solution and the attachment kernel is  $\Theta_i = X_i/t = c_i$ . Substituting into Eq. (3), we obtain the recurrence relation  $c_{i+1} = c_i/2$ , which yields the familiar degree distribution  $c_i = 2^{-i}$ . Substituting into Eq. (4) yields the recurrence relation

$$\begin{aligned} l_{i,j} &= (l_{i-1,j} + l_{i,j-1})/3, \quad i, j > 1, \\ l_{1,j} &= (c_{j-1} + l_{1,j-1})/3, \quad j > 1, \end{aligned} \quad (10)$$

with  $l_{1,1} = 0$ . The exact solution for  $l_{i,j}$  is

$$\begin{aligned} l_{i,j} &= \sum_{x=2}^j \frac{\binom{i-1+j-x}{j-x}}{3^{(i+j-x)}2^{(x-1)}} + \sum_{x=2}^i \frac{\binom{i-1+j-x}{i-x}}{3^{(i+j-x)}2^{(x-1)}}, \quad i, j > 1, \\ l_{1,j} &= \sum_{k=1}^{j-1} (3^k 2^{j-k})^{-1}, \quad j > 1, \end{aligned} \quad (11)$$

where  $\binom{x}{y}$  is the conventional combinatorial ‘choose’ function. Further mathematical details are given in Appendix A. We shall make use of this solution to investigate the correlations of such a network in Section III. This normalized link-space matrix is illustrated in Fig. 1 along with a comparison to simulated networks using the cumulative link-space matrix.

### B. Barabási-Albert (BA) model (steady-state solution).

In the BA model [8, 9], at each timestep a new node is added to the network and connected to  $m$  existing nodes with probabilities proportional to the degrees of those nodes, i.e.  $\Theta_i \propto i$  yielding

$$\Theta_i(t) = \frac{iX_i(t)}{\sum_j jX_j(t)} = \frac{iX_i(t)}{2M(t)} \approx \frac{iX_i(t)}{2mt}. \quad (12)$$

We consider the scenario when the new node is added with one undirected link,  $m = 1$ , and the attachment kernel is well approximated by  $\Theta_i \approx ic_i/2$ . Substituting this into Eq. (3) yields the recurrence relation  $c_i = \frac{(i-1)c_{i-1}}{2} - \frac{ic_i}{2} = \frac{i-1}{i+2}c_{i-1}$ , whose solution is  $c_i = \frac{4}{i(i+1)(i+2)}$ . Using the same substitution, the link-space master equations yield the recurrence relations:

$$\begin{aligned} l_{i,j} &= \frac{(i-1)l_{i-1,j} + (j-1)l_{i,j-1}}{2+i+j}, \quad i, j > 1, \\ l_{1,j} &= \frac{(j-1)c_{j-1} + (j-1)l_{1,j-1}}{3+j}, \quad j > 1. \end{aligned} \quad (13)$$

The exact solution for  $l_{i,j}$  is obtained by algebraic manipulation of the normalised link-space matrix using the previously derived degree distribution (further mathematical details are given in Appendix B) and is given by

$$\begin{aligned} l_{i,j} &= \frac{4(j-1)!(i-1)!}{(j+i+2)!} \left\{ G(i+1) + 2G(i) - 3G(i-1) \right. \\ &\quad \left. + \frac{1}{2} \sum_{x=2}^i (x-1)(x+6) [G(i-x) - G(i-x-1)] \right\}, \end{aligned} \quad (14)$$

$$\text{where } G(x) = \begin{cases} \frac{(j+x-1)!}{x!(j-1)!} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases}$$

The first few rows of this matrix have the form:

$$\begin{aligned} l_{1,j} &= \frac{2(j+6)(j-1)}{j(j+1)(j+2)(j+3)}, \\ l_{2,j} &= \frac{2j(j-1)(j+10) + 48}{3j(j+1)(j+2)(j+3)(j+4)}, \\ l_{3,j} &= \frac{3j^4 + 42j^3 - 3j^2 + 246j + 360}{9j(j+1)(j+2)(j+3)(j+4)(j+5)}. \end{aligned} \quad (15)$$

We shall use this solution to investigate the correlations with this network in Section III. This normalized link-space matrix is illustrated in Fig. 2 along with a comparison to simulated networks.

### C. Erdős and Rényi random graph (time-dependent solution).

The Erdős and Rényi (ER) classical random graph is the quintessential equilibrium network model [2, 10]. However, to employ the link-space formalism, which tracks the evolution of a network’s correlation properties, we must model it as an evolving, non-equilibrium network. To do so is a straightforward process and the model proceeds as follows. At each timestep, we add one new node to the existing network. All possible links between the new node and *all* existing nodes are considered and each is established with probability  $\alpha$ . That is, a biased coin toss (Bernoulli trial) is employed for every node within the existing network to decide whether a link is formed between it and the new node. Consequently the expected number of new, undirected links with which the new node connects to the existing network is  $\langle m(t) \rangle = \alpha N(t-1) \approx \alpha t$  and this would be described as an accelerating network [36]. No new links are formed between existing nodes. This model subsequently produces a network in which the probability of a link existing between any pair of nodes is simply  $\alpha$  and is thus representative of the ER random graph [10]. A network grown to  $t$  nodes will have mean degree  $\alpha t$  which will also be the expected degree of all nodes in the network. Whilst the random attachment and preferential attachment models both have steady-state asymptotic behaviour, clearly this model doesn’t and a full, time-dependent solution is required.

Using the random attachment probability kernel  $\Theta_i(t) = X_i(t)/N(t)$ , the node-space master equation for the number of nodes of degree  $i$  can be written

$$\begin{aligned} \langle X_i(t+1) \rangle &= X_i(t) + \frac{\langle m(t) \rangle X_{i-1}(t)}{N(t)} - \frac{\langle m(t) \rangle X_i(t)}{N(t)} \\ &\quad + P\{m(t) = i\}. \end{aligned} \quad (16)$$

The second term on the right hand side reflects the expected number of connections to  $i-1$  degree nodes making them nodes of degree  $i$ . The last term is the probability that the new node itself is a node of degree  $i$ . This will be binomially distributed and, if we assume that  $t$  is large, we can make a Poisson approximation. Recalling that the fraction of nodes of degree  $i$  is given by  $c_i(t) = X_i(t)/N(t) = X_i(t)/t$ , we can rewrite Eq. (16) as

$$\frac{d(tc_i(t))}{dt} = \alpha c_{i-1}(t) - \alpha c_i(t) + \frac{e^{-\alpha t}(\alpha t)^i}{i!}. \quad (17)$$

The time-dependent solution of Eq. (17) yields the degree distribution for this model and is simply

$$c_i(t) = \frac{e^{-\alpha t}(\alpha t)^i}{i!}, \quad (18)$$

which is what we would expect for the random graph with mean degree of  $\alpha t$  in the large size limit [37]. We can similarly write the link-space master equation for this

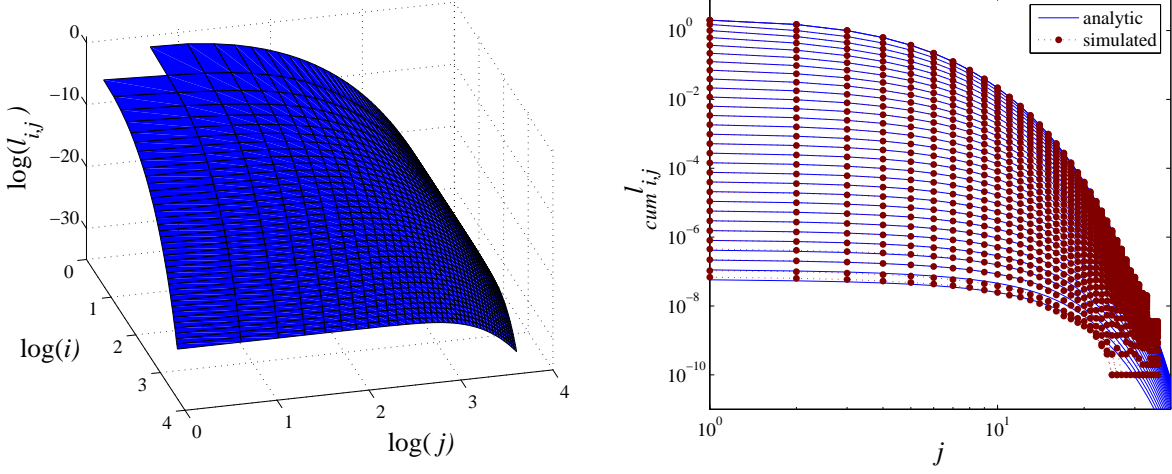


FIG. 1: (Color Online) Left: the analytically-derived, normalized link-space matrix for the random attachment growth algorithm. This could be filled to arbitrary size but is truncated to maximum degree of 40 here. Right: comparison of the cumulative link-space matrices for the analytic solution and a simulation of the algorithm. The simulation comprises an ensemble average of 100 networks grown to  $10^8$  nodes. The maximum degree obtained was 37. The first 30 rows ( $i$  values) of the cumulative link-space are illustrated and finite-size effects are noticeable at high  $j$ .

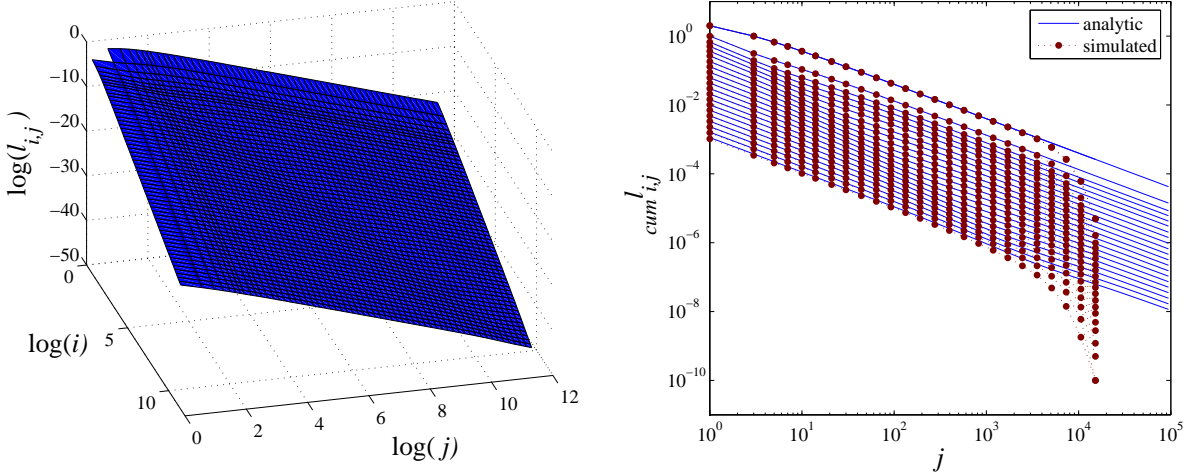


FIG. 2: (Color Online) Left: the analytically-derived, normalized link-space matrix for the preferential attachment growth algorithm. This could be filled to arbitrary size but is here truncated to maximum degree of approximately  $10^5$ . Right: comparison of the cumulative link-space matrices for the analytic solution and a simulation of the algorithm. The simulation comprises an ensemble average of 1000 networks each grown to  $10^7$  nodes. The maximum degree obtained was 17609. Various rows ( $i$  values) from 1 to 3500 of the cumulative link-space are illustrated and the effects of the finite nature of the simulation are apparent.

model:

$$\begin{aligned}
 \langle L_{i,j}(t+1) \rangle &= L_{i,j}(t) + \frac{\langle m(t) \rangle \Theta_{i-1}(t) L_{i-1,j}}{X_{i-1}} \\
 &+ \frac{\langle m(t) \rangle \Theta_{j-1}(t) L_{i,j-1}}{X_{j-1}} - \frac{\langle m(t) \rangle \Theta_i(t) L_{i,j}}{X_i} \\
 &- \frac{\langle m(t) \rangle \Theta_j(t) L_{i,j}}{X_j} + i \Theta_{j-1} P\{m(t) = i\} \\
 &+ j \Theta_{i-1} P\{m(t) = j\}.
 \end{aligned} \tag{19}$$

The last two terms reflect the expected number of new  $i \leftrightarrow j$  links being formed between the new node and existing nodes. Recalling that the normalized link-space matrix is given by  $l_{i,j}(t) = L_{i,j}(t)/M(t)$  where the number of links,  $M(t)$ , at time  $t$  can be approximated as  $M(t) \approx \alpha t^2/2$  and making use of the previously derived, time-dependent degree distribution such that  $P\{m(t) =$

$j\} = c_j(t)$ , Eq. (19) can be rewritten as

$$\frac{d}{dt} \left( \frac{\alpha t^2 l_{i,j}(t)}{2} \right) = \frac{(\alpha t)^2}{2} (l_{i-1,j}(t) + l_{i,j-1}(t) - 2l_{i,j}(t)) + \frac{2ijc_i(t)c_j(t)}{\alpha t}. \quad (20)$$

The solution to Eq. (20) is found by means of ansatz (the choice of which is explained in Section III) and is

$$l_{i,j}(t) = \frac{2e^{-2\alpha t}(\alpha t)^{i+j-2}}{(i-1)!(j-1)!}. \quad (21)$$

This is the normalized link-space matrix for a classical random graph with a mean degree of  $\alpha t$  in the large size limit and this is illustrated in Fig 3 where a comparison to a simulation is also made.

### III. DEGREE CORRELATIONS AND ASSORTATIVITY

The average nearest-neighbor degree  $\langle k_{nn} \rangle_i$  of nodes of degree  $i$  can be easily obtained from the link-space matrix and is given by

$$\langle k_{nn} \rangle_i = \frac{\sum_{j=1}^{\infty} j L_{i,j}}{\sum_{j=1}^{\infty} L_{i,j}} = \frac{\sum_{j=1}^{\infty} j l_{i,j}}{\sum_{j=1}^{\infty} l_{i,j}}. \quad (22)$$

This is illustrated in Fig. 4 for the steady-state solutions of the random attachment model, the BA model and the ER random graph. If the average nearest-neighbour degree  $\langle k_{nn} \rangle_i$  is constant with respect to degree  $i$ , non-assortativity is implied. Certainly, the random attachment curve continues to increase, implying positive assortative mixing. The preferential attachment curve in Fig. 4 appears to asymptote to a constant value (as was noted in [26]) although the fact that it decreases initially illustrates that it is not perfectly non-assortative. This is a feature that would not be captured using a normalised Pearson's correlation coefficient which would suggest non-assortativity for this network [3, 38].

The link-space formalism allows us to address two-vertex correlations in a more powerful way in that we can calculate the conditional, two-point vertex degree distribution,  $P(j|i)$  which has been traditionally difficult to measure [19]. We can write this joint probability, i.e. the probability that a randomly chosen edge is connected to a node of degree  $j$  given that the other end is connected to a node of degree  $i$ , in terms of the link-space matrix (normalized or not) as  $P(j|i) = L_{i,j}/(iX_i) = L_{i,j}/\sum_{j=1}^{\infty} L_{i,j} = l_{i,j}/\sum_{j=1}^{\infty} l_{i,j}$ . For total number of edges approximately equal to the total number of nodes, this can be approximated as  $l_{i,j}/(ic_i)$ . This is illustrated in Fig. 5 for the scenarios of random and preferential attachment where the non-trivial nature of the correlations present within these networks is evident.

Consider selecting an edge at random in the network. If we then select one end of this edge at random, the probability that this node has degree  $j$  will be proportional

to  $j$  since higher degree nodes have, by definition, more links connected to them than low degree nodes. Consider now only a subset of all edges with one end attached to a node of degree  $i$ . If there were no correlations present, the probability that the other end is attached to a node of degree  $j$  is again proportional to  $j$  [39]. The criterion of perfect non-assortativity for a network with equal numbers of nodes and links can be described as

$$P(j|i) = \frac{jc_j}{2}, \quad \forall j. \quad (23)$$

This criterion is illustrated in Fig. 5 for the random attachment and BA models and confirms that neither is perfectly non-assortative.

#### A. Perfect non-assortativity

It is interesting to ask whether a perfectly non-assortative network can be generated. Assuming that such a network will not have equal numbers of nodes and links, we must rewrite the conditions of non-assortativity accordingly for total number of edges  $M$ :

$$P(j|i) = P(j) = \frac{jX_j}{2M} = \frac{jNc_j}{2M}. \quad (24)$$

Recalling our definition of the normalized link-space matrix that  $l_{i,j} = L_{i,j}/M$ , we can express this conditional probability  $P(j|i)$  in terms of this matrix as

$$P(j|i) = \frac{L_{i,j}}{iX_i} = \frac{Ml_{i,j}}{iNc_i}. \quad (25)$$

As such, we can now write for  $l_{i,j}$

$$l_{i,j} = \left( \frac{N}{M} \right)^2 \frac{ic_i jc_j}{2}. \quad (26)$$

That is, for *any* degree distribution, a normalized link-space matrix can be found which is representative of a perfectly non-assortative network [48]. This might provide an alternative to the network randomization technique of Maslov and Sneppen to provide a “null model network” [32]. The expression of Eq. (26) provides the ansatz solution for the normalized link-space matrix to the master equations for the ER random graph in Section II C.

### IV. DECAYING NETWORKS

Although somewhat counter-intuitive, it is possible to find steady states of networks whereby nodes and/or links are removed from the system. Aside from the obvious situation of having no nodes or edges left, we would like to investigate the possible existence of a network configuration whose link-space matrix and, subsequently, node degree distribution are static with respect to the decay



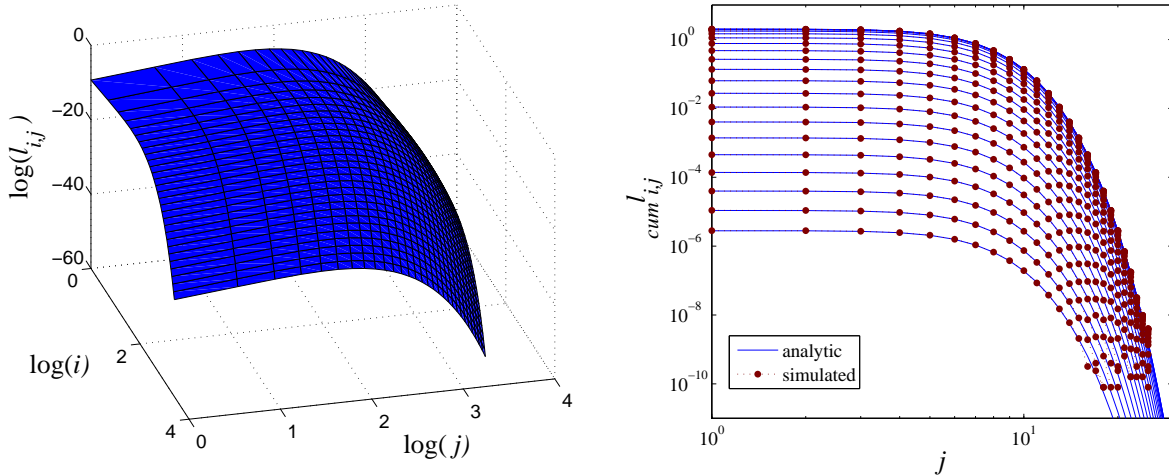


FIG. 3: (Color Online) Left: the analytically-derived, normalized link-space matrix for the Erdős and Rényi (ER) random graph with mean degree equal to 5. This could be filled to arbitrary size but is here truncated to maximum degree of 30. Right: comparison of the cumulative link-space matrices for the analytic solution and a simulation of the algorithm. The simulation comprises an ensemble average of 500 networks each consisting of  $10^7$  nodes. The probability that a link exists between any pair of nodes is  $\alpha = 5 \times 10^{-7}$  and the maximum degree obtained was 25. The first 20 rows ( $i$  values) of the cumulative link-space are illustrated and, again, the finite nature of the simulation is apparent for high  $j$ .

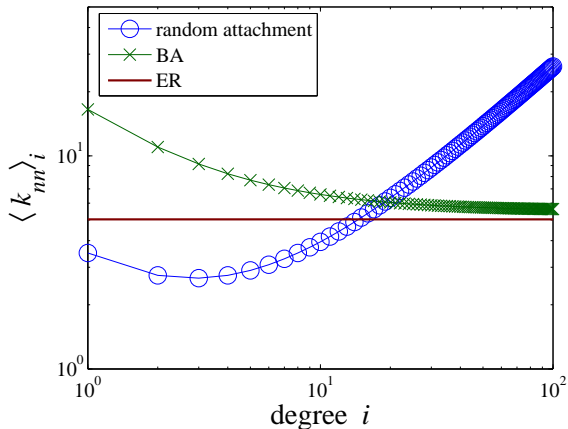


FIG. 4: (Color Online) The mean degree of the nearest neighbours  $\langle k_{nn} \rangle_i$  of nodes of degree  $i$  as a function of  $i$  for the analytic solutions to the random attachment algorithm (Section II A), the BA model (Section II B) and the ER random graph (Section II C).

process. The concept that decay processes are highly influential on a network's structure has been considered before although this has typically only been investigated in conjunction with simultaneous growth [40, 41, 42]. Here we shall employ the link-space formalism to examine the effect of some simple, decay-only scenarios, specifically the two simplest cases – random link removal and random node removal.

#### A. Random Link Removal (RLR)

Consider an arbitrary network. At each timestep, we select a fixed number,  $w$ , of links at random and remove them. We shall implement the link-space formalism to investigate whether or not it is possible that such a mechanism can lead to stationary structure. Consider the link-space element  $L_{i,j}(t)$  denoting the number of links from nodes of degree  $i$  to nodes of degree  $j$ . Clearly this can be decreased if an  $i \leftrightarrow j$  link is removed, i.e. a link that connects a degree  $i$  node to a degree  $j$  node. Also, if a  $k \leftrightarrow i$  link is removed and that  $i$  node has further links to  $j$  degree nodes, then those that were  $i \leftrightarrow j$  links will now become  $(i-1) \leftrightarrow j$ , similarly for  $k \leftrightarrow j$  links being removed. However, if the link removed is a  $k \leftrightarrow (j+1)$  link and that  $j+1$  degree node is connected to a degree  $i$  node, then when the  $j+1$  node becomes a degree  $j$  node, the  $(j+1) \leftrightarrow i$  link will become an  $j \leftrightarrow i$  link, increasing  $L_{i,j}$ . Let us assume that we are removing links at random from the network comprising  $N(t)$  vertices and  $M(t)$  links. A non-random link selection process could be incorporated into the master equations using a probability kernel. The master equation for this process can be written in terms of the expected increasing and decreasing

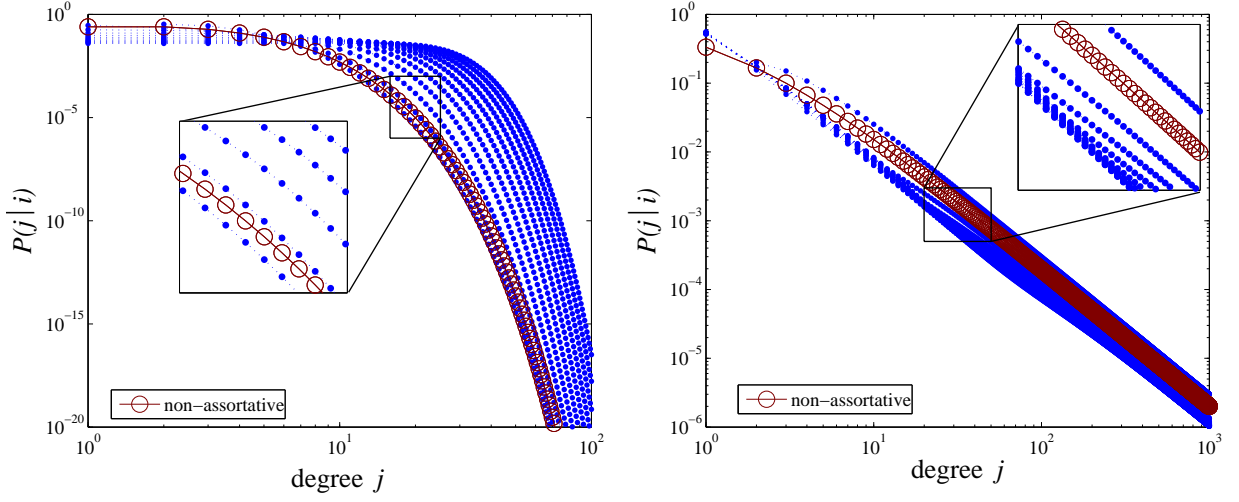


FIG. 5: (Color Online) The conditional probability  $P(j|i)$  that one end of a randomly selected link has degree  $j$  given that the other end is of degree  $i$  over the range  $i = \{1, 6, 11, 16, \dots, 51\}$ . The left plot is for the random attachment algorithm and the right is preferential attachment. The criterion of non-assortativity of Eq. (23) is denoted with red circles.

contributions:

$$\begin{aligned}
 \langle L_{i,j}(t+1) \rangle &= L_{i,j}(t) + w \sum_k \frac{L_{k,j+1}(t)jL_{i,j+1}(t)}{M(t)(j+1)X_{j+1}(t)} \\
 &+ w \sum_k \frac{L_{k,i+1}(t)iL_{i+1,j}(t)}{M(t)(i+1)X_{i+1}(t)} \\
 &- w \sum_k \frac{L_{k,j}(t)(j-1)L_{i,j}(t)}{M(t)(j)X_j(t)} \\
 &- w \sum_k \frac{L_{k,i}(t)(i-1)L_{i,j}(t)}{M(t)(i)X_i(t)} \\
 &- \frac{wL_{i,j}(t)}{M(t)}. \tag{27}
 \end{aligned}$$

This simplifies to

$$\begin{aligned}
 \langle L_{i,j}(t+1) \rangle &= L_{i,j}(t) + \frac{wjL_{i,j+1}(t)}{M(t)} \\
 &+ \frac{wiL_{i+1,j}(t)}{M(t)} - \frac{w(i-1)L_{i,j}(t)}{M(t)} \\
 &- \frac{w(j-1)L_{i,j}(t)}{M(t)} - \frac{wL_{i,j}(t)}{M(t)}, \tag{28}
 \end{aligned}$$

where the last term refers to the physical removal of an  $i \leftrightarrow j$  link.

To investigate the possibility of a steady state solution, we make a similar argument as before for growing networks but this time for the process of removing  $w$  links per timestep:

$$\begin{aligned}
 L_{i,j}(t) &= l_{i,j}M(t) \\
 &= l_{i,j}(M_0 - wt), \\
 \frac{dL_{i,j}}{dt} &= -wl_{i,j} \\
 &\approx L_{i,j}(t+1) - L_{i,j}(t). \tag{29}
 \end{aligned}$$

The expression of Eq. (28) can be reduced to

$$l_{i,j} = \frac{il_{i+1,j} + jl_{i,j+1}}{i+j-2}. \tag{30}$$

As such, (twice) the fraction of  $1 \leftrightarrow 1$  links,  $l_{1,1}$ , does not reach a steady state. This might be expected, as the removal process for such links requires them to be physically removed as opposed to the process by which the degree of the node at one end of the link is being decreased. Therefore, the value  $l_{1,1}$  increases in time. However, we can investigate the properties of the rest of the links in the network which do reach a steady state by neglecting these two node components (see Appendix C for details). The normalized link-space matrix can be written for this system for  $i+j \neq 2$  as

$$l_{i,j} = \frac{A}{2^{i+j}} \frac{(i+j-3)!}{(i-1)!(j-1)!}. \tag{31}$$

The summation over this link-space matrix does not converge, i.e.  $\sum_i \sum_j l_{i,j} \rightarrow \infty$ . However, we can still derive the degree distribution using Eq. (5) and normalize this such that  $\sum_i c_i = 1$  (see Appendix C for details):

$$\begin{aligned}
 c_1 &= \frac{\log(2)}{1 + \log(2)}, \\
 c_i &= \frac{1}{(1 + \log(2))i(i-1)} \quad i > 1. \tag{32}
 \end{aligned}$$

Consequently, the first moment of the distribution diverges also,  $\sum_i ic_i = \langle k \rangle \rightarrow \infty$ . It is interesting that the process of randomly removing links generates a network which exhibits power-law scaling of exponent 2, corresponding to a degree distribution significantly over-skewed with respect to the preferential selection process

of the BA growth algorithm (exponent 3). By considering the average degree of the neighbours of nodes of degree  $i$ , as discussed in Section III, we can see that the random link removal algorithm generates highly assortative networks.

### B. Random Node Removal (RNR)

In a similar manner to Section IV A, we will now discuss the possibility of creating such a steady-state structure via a process of removing nodes (with all of their links) from an existing network at the rate of  $w$  nodes per timestep. Clearly, removing some node which has a link to an  $(i+1)$ -degree node which in turn has a link to a  $j$ -degree node can increase the number of  $i \leftrightarrow j$  links in the system. The other processes which can increase or decrease the number of links from  $i$  degree nodes to  $j$  degree nodes can be similarly explained. We consider selecting a node of some degree  $k$  for removal with some probability kernel  $\Theta_k$  (as in the growing algorithms of Section II). The master equation for such a process can thus be written for general node selection kernel:

$$\begin{aligned} \langle L_{i,j}(t+1) \rangle = & L_{i,j}(t) - w \frac{\Theta_i(t)L_{i,j}(t)}{X_i(t)} - w \frac{\Theta_j(t)L_{i,j}(t)}{X_j(t)} \\ & + w \sum_k \frac{\Theta_k(t)}{X_k(t)} \left\{ L_{k,i+1}(t) \frac{iL_{i+1,j}(t)}{(i+1)X_{i+1}(t)} \right. \\ & \left. + L_{k,j+1}(t) \frac{jL_{i,j+1}(t)}{(j+1)X_{j+1}(t)} \right\} \\ & - w \sum_k \frac{\Theta_k(t)}{X_k(t)} \left\{ L_{k,i}(t) \frac{(i-1)L_{i,j}(t)}{iX_i(t)} \right. \\ & \left. + L_{k,j}(t) \frac{(j-1)L_{i,j}(t)}{jX_j(t)} \right\}. \end{aligned} \quad (33)$$

We now specify a kernel for selecting the node to be removed, namely, the random kernel although the approach is general to any node selection procedure. Selecting a node purely at random leads to  $\Theta_k = c_k$ . The steady-state assumptions must be clarified slightly. As before

$$L_{i,j}(t) = l_{i,j}M(t). \quad (34)$$

However, because we can remove more than one link (through removing a high degree node for example) we must approximate for  $M(t)$ . Using the random removal kernel, we can assume that on average, the selected node will have degree equal to the mean degree of the network,  $\langle k(t) \rangle = \frac{2M(t)}{N(t)}$ . We can use this to write the number of

remaining links in the network as

$$\begin{aligned} L_{i,j}(t) &= l_{i,j}(M(0) - w\langle k(t) \rangle) \\ &= l_{i,j}(M_0 - \frac{2wM(t)}{N(t)}t), \\ \frac{dL_{i,j}}{dt} &= -\frac{2wM(t)}{N(t)}l_{i,j} \\ &\approx L_{i,j}(t+1) - L_{i,j}(t). \end{aligned} \quad (35)$$

Making use of the link-space identities and Eq. (35), the master equation Eq. (33) can be written in simple form for the steady state as

$$l_{i,j} = \frac{il_{i+1,j} + jl_{i,j+1}}{i+j-2}. \quad (36)$$

Clearly, this is identical to Eq. (30) for the random link removal model of Section IV A and, consequently, the analysis of the degree distribution will be the same too. It is interesting that the random node removal and random link removal lead to the same degree correlations.

### V. MIXTURE MODEL

In this section we introduce a simple model that makes use of only local information about node degrees as microscopic mechanisms requiring global information are often unrealistic for many real-world networks [43]. It therefore provides insight into possible alternative microscopic mechanisms for a range of biological and social networks. The link-space formalism allows us to identify the transition point at which lower power-law exponent degree distributions switch to higher exponent distributions with respect to the BA preferential attachment model. While similar local algorithms have been proposed in the literature [11, 44, 45], the strength of the approach followed here is the ability to describe the inherent degree-degree correlations.

It is well known that a mixture of random and preferential attachment in a growth algorithm can produce power law-degree distributions with exponents  $\gamma \in [3, \infty)$  when new links are only established between the new node and the existing network [49]. It has often been assumed that a *one step* random walk replicates linear preferential attachment [43, 45]. This is *not* true. A one step random walk is in fact more biased towards high degree nodes than preferential attachment [50] as can be easily seen by performing the procedure on a simple hub and spoke network. In this case, the probability of arriving at the hub tends to one for increasingly large networks rather than a half as would be appropriate for preferential attachment. We can use this bias to generate networks with degree distributions that have lower power-law exponent than the preferential attachment model, a phenomenon also observed by Evans and Saramäki [44]. A mixture of this approach with random attachment results in a simple model that can span a wide range of degree distributions. This one-parameter, growing network growth

model, in which we simply attach a single node at each timestep with a single undirected link, does not require prior knowledge of the existing global network structure.

The algorithm proceeds explicitly as follows: (i) pick a node  $\kappa$  within the existing network at random; (ii) with probability  $a$  make a link to that node; otherwise (iii) pick any of the neighbours of  $\kappa$  at random and link to that node. Hence this algorithm resembles an object or ‘agent’ making a short random-walk. This is very similar to the Growing Network with Redirection model introduced by Krapivsky and Redner [11] except that the model here employs undirected links which necessitates a random (as opposed to deterministic) walk. Fig. 6 shows examples of the resulting networks and the corresponding (cumulative) degree distributions. Interestingly,  $a = 0$  yields a network that is dominated by hubs and spokes while  $a = 1$  yields the random attachment network. Intermediate values of  $a$  yield networks which are neither too ordered nor too disordered. For  $a \sim 0.2$ , the algorithm generates networks whose degree-distribution resembles the BA preferential-attachment network (see Figs. 6 and 7).

Our analysis of the algorithm starts by establishing the attachment kernel,  $\Theta_i(t)$ , which in turn requires properly resolving the one-step random walk. The link-space formalism provides us with an expression for the probability  $P'_i(t)$  associated with performing a random walk of length one and arriving at a degree  $i$  node [38]:

$$P'_i(t) = \frac{X_1(t)L_{1,i}(t)}{N(t)X_1(t)} + \frac{X_2(t)L_{2,i}(t)}{2N(t)X_2(t)} + \frac{X_3(t)L_{3,i}(t)}{3N(t)X_3(t)} + \dots \quad (37)$$

This can also be written [38] as  $P'_i(t) = \frac{iX_i(t)}{N(t)} \langle \frac{1}{k_{nn}(t)} \rangle_i$  where the average is performed over the neighbours of nodes with degree  $i$ . Note that this quantity does *not* replicate preferential attachment, in contrast to what is commonly thought [43, 45]. Defining  $\beta_i(t)$  as

$$\beta_i(t) \equiv \frac{1}{i} \frac{1}{c_i(t)} \langle \frac{1}{k_{nn}(t)} \rangle_i = \frac{\sum_k \frac{L_{i,k}(t)}{k}}{\sum_k L_{i,k}(t)} = \frac{\sum_k \frac{l_{i,k}(t)}{k}}{\sum_k l_{i,k}(t)}, \quad (38)$$

yields

$$\Theta_i(t) = ac_i(t) + (1-a)\beta_i(t)ic_i(t). \quad (39)$$

To investigate a possible steady-state, we could substitute the above equations into Eqs. (3) and (4). However the non-linear terms resulting from  $\beta$  imply that a complete analytical solution for  $l_{i,j}$  and  $c_i$  is difficult. We leave this as a future challenge, but stress that our formalism can be implemented in its non-stationary form Eq. (2) numerically by iteration with very good efficiency [38], yielding the degree distributions shown in Fig. 7.

We can now use the link-space formalism to deduce the parameter value at which our algorithm yields the BA degree distribution in the long-time limit. At this value  $a = a_c$ , the node-degree distribution goes from lower to higher power-law exponents with respect to the BA

model. For this parameter value, the attachment probability to nodes of various degrees is equal for both our mixture algorithm and the BA model. Using Eqs. (12) and (39), we have  $\frac{i}{2} = a_c + (1-a_c)\beta_i i$ , and hence for large  $i$  this yields  $a_c = 1 - \frac{1}{2\beta_i}$ . We could then proceed to use the exact solution of the link-space equations for the preferential-attachment algorithm, in order to infer  $\beta_i$  in the high  $i$  limit. However, since  $\beta_i$  can be expanded in terms of  $l_{i,j}$  as shown in Eq. (38) and  $l_{i,j}$  decays very rapidly as  $i, j$  become large, we can obtain a good approximation by using only the first two terms of Eq. (15). This yields  $\beta \approx 0.66$ . Hence, the critical value at which this simple model approximates the BA degree distribution is  $a_c = 0.25$ , illustrated with the inset plot of Fig. 7.

## VI. CONCLUSION

We have developed a new formalism which accounts for degree-degree correlations in networks. We have employed the formalism to produce analytic solutions to the link-space matrix for the random attachment, BA and ER network models allowing detailed analysis of the degree correlations therein. We have introduced a cumulative implementation of the formalism which can be used to compare link-space matrices and might also be applied to empirical networks. We have shown that a perfectly non-assortative network can be generated with arbitrary degree distribution. We have also demonstrated the possibility of a steady-state of the degree distribution and of the two-point degree correlations for simple decaying networks through deriving the static, normalized link-space matrix for such a process. Employing the link-space formalism allowed us to accurately describe a simple one-parameter network growth algorithm which is able to reproduce a wide variety of degree distributions without any global information about node degrees. We have used the framework to show that a one-step random walk does *not* replicate preferential attachment except in the particular case of a perfectly non-assortative network. Indeed this in itself represents a criterion for perfect non-assortativity. Whilst the present paper focuses on introducing and illustrating the use of the link-space formalism for a variety of model networks, we stress that its applicability is far more general.

**Acknowledgements:** For interesting and helpful suggestions, the authors would like to thank T.S. Evans, M.A. Porter and one anonymous referee. D.M.D.S. is supported by the European Union (MMCOMNET), C.F.L. by a Glasstone Fellowship (Oxford), and J-P.O. by a Wolfson College Junior Research Fellowship (Oxford).

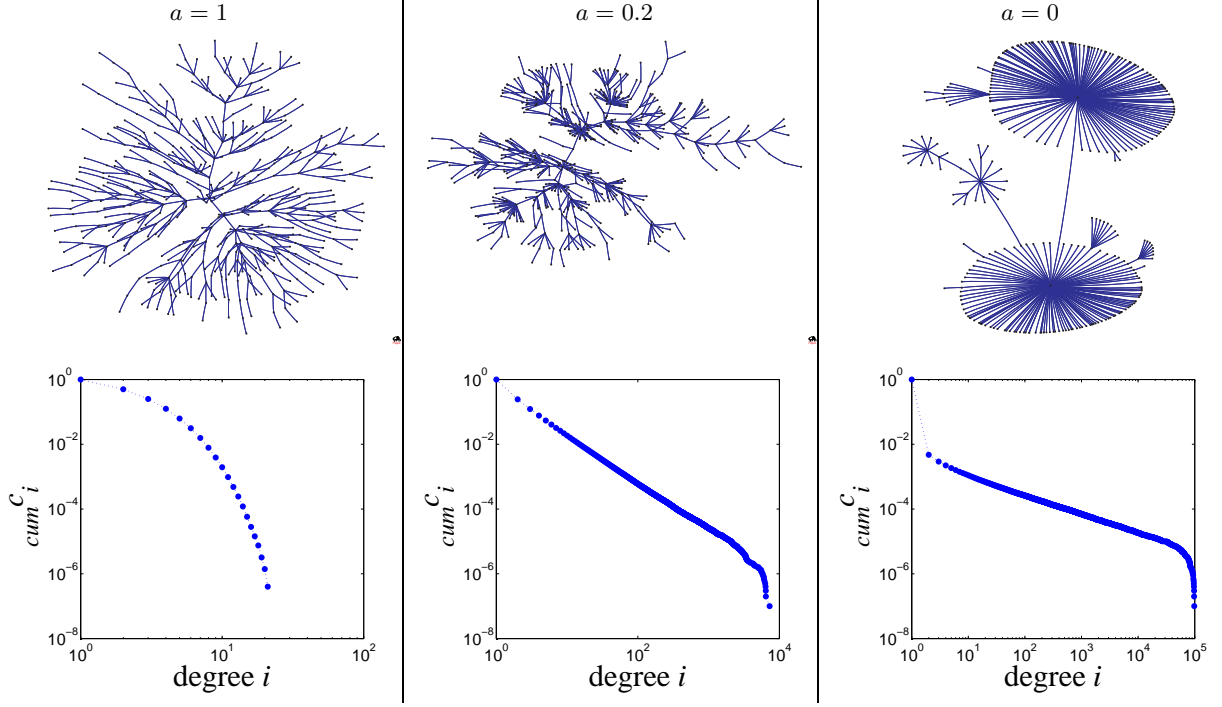


FIG. 6: (Color Online) Top: Networks generated using the one-parameter, local information growth algorithm with  $a = 1$  (left),  $a = 0.2$  (center),  $a = 0$  (right). Initial network seed comprised two nodes and one link. Networks drawn with Pajek [46]. Bottom: Cumulative degree distributions for the same networks grown to  $10^5$  nodes and ensemble-averaged over 100 networks per  $a$  value.

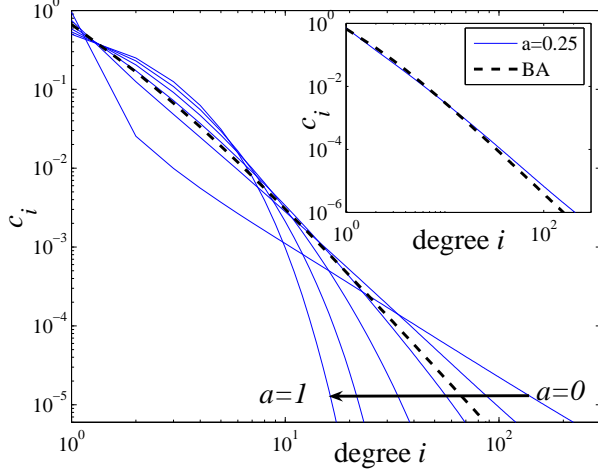


FIG. 7: (Color Online) Degree distributions generated using the link-space analysis of our one-step algorithm. Dashed line: results for the Barabási-Albert (BA) [8] preferential-attachment algorithm. Our one-step algorithm is closely resembles the BA results with parameter value  $a = 0.25$  (see inset).

## APPENDIX A: EXACT SOLUTION OF RANDOM ATTACHMENT

For the random attachment algorithm, where one new node is added to the existing network with one new link at each timestep, the link-space master equation can be expressed for the steady state as

$$\begin{aligned} l_{i,j} &= \frac{l_{i-1,j} + l_{i,j-1}}{3}, \quad i, j > 1, \\ l_{1,j} &= \frac{c_{j-1} + l_{1,j-1}}{3}, \quad j > 1, \\ l_{1,1} &= 0. \end{aligned} \quad (\text{A1})$$

It is easy to populate the link-space matrix numerically, just from the degree distribution  $c_i = 2^{-i}$  obtained from solving the node-space recurrence relation and from the link-space master equation Eq. (A1). At first glance, the solution to the master equation A1 would be of the form (check by substitution)

$$l_{i,j} = \frac{b}{4 \cdot 2^{i+j} 3^{i+j}}. \quad (\text{A2})$$

However, the boundary conditions, which could be interpreted as influx of probability into the diffusive matrix, are such that this does not hold. We can actually solve the normalized link-space matrix for this model exactly. Consider the values  $c_i$  as being influxes of probability

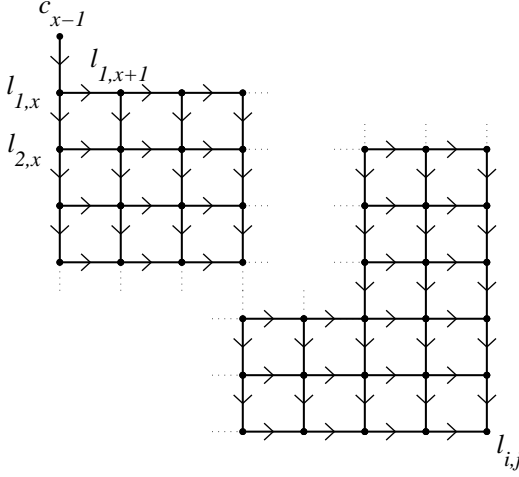


FIG. 8: The paths of probability flux from  $c_{x-1}$  influencing element  $l_{i,j}$ . Each arrow (step) represents a further factor of  $\frac{1}{3}$ .

into the top and left of the link-space matrix. We can compute the effect of one such element on the value in the matrix  $l_{i,j}$ . Each step in the path of probability flux reflects an extra factor of  $\frac{1}{3}$ . First, we will consider the influx effect from the top of the matrix as in Fig. 8. The total path length from influx  $c_{x-1}$  to element  $l_{i,j}$  is simply  $i + j - x$ . The number of possible paths between  $c_{x-1}$  and element  $l_{i,j}$  can be expressed as  $\binom{j-x+i-1}{j-x-1}$ . We can similarly write down the paths and lengths for influxes into the left hand side of the matrix. The first row (and column) can be described as

$$\begin{aligned} l_{1,j} &= \frac{c_{j-1} + l_{1,j-1}}{3} \\ &= \frac{2^{-(j-1)} + l_{1,j-1}}{3} \\ &= \sum_{k=1}^{j-1} \frac{1}{3^k 2^{j-k}}. \end{aligned} \quad (\text{A3})$$

Subsequent rows can be similarly described. So, for  $i, j > 1$  an element can be written:

$$\begin{aligned} l_{i,j} &= \frac{l_{i-1,j} + l_{i,j-1}}{3} \\ &= \sum_{x=2}^j \frac{\binom{i-1+j-x}{j-x}}{3^{(i+j-x)} 2^{(x-1)}} + \sum_{x=2}^i \frac{\binom{i-1+j-x}{i-x}}{3^{(i+j-x)} 2^{(x-1)}}. \end{aligned} \quad (\text{A4})$$

The results can be observed in Fig. 9 where a comparison to the numerically populated link-space matrix is made.

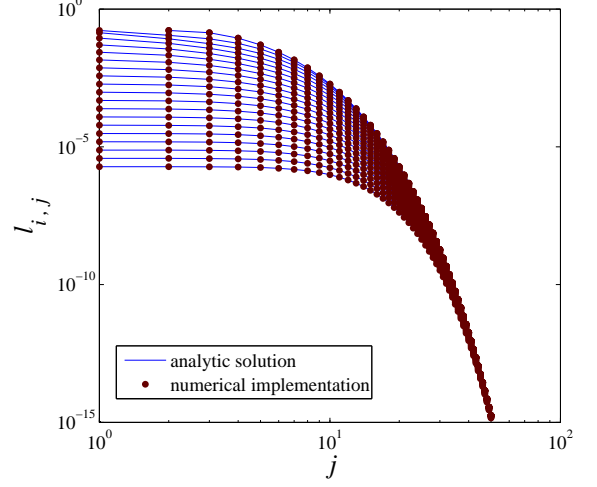


FIG. 9: (Color Online) Comparison of the numerically derived link-space matrix (markers) and the analytic solution (solid lines) for random node attachment for the first 20 rows ( $i$  values) of the link-space matrix. The numerical implementation populates the link-space matrix directly from the degree distribution and the steady-state link-space recurrence relations.

## APPENDIX B: EXACT SOLUTION OF PREFERENTIAL ATTACHMENT

In the BA model, when adding one new link with one undirected node per timestep, the steady-state solution of the node-space master equation leads to the degree distribution

$$c_i = \frac{4}{i(i+1)(i+2)}, \quad (\text{B1})$$

which can be checked easily by substitution. Substituting the attachment probability kernel  $\Theta_i \approx ic_i/2$  into the link-space master equation Eq. (4), we obtain the master equations for the link-space for this network growth algorithm:

$$\begin{aligned} l_{i,j} &= \frac{(i-1)l_{i-1,j} + (j-1)l_{i,j-1}}{2+i+j}, \quad i, j > 1, \\ l_{1,j} &= \frac{(j-1)c_{j-1} + (j-1)l_{1,j-1}}{3+j}, \quad j > 1, \\ l_{1,1} &= 0. \end{aligned} \quad (\text{B2})$$

Again, it is easy to populate the matrix numerically just by implementing the node and link-space equations. At first glance, the solution to the master equation Eq. (B2) would be of the form (check by substitution)

$$l_{i,j} = \frac{A}{i(i+1)j(j+1)}, \quad (\text{B3})$$

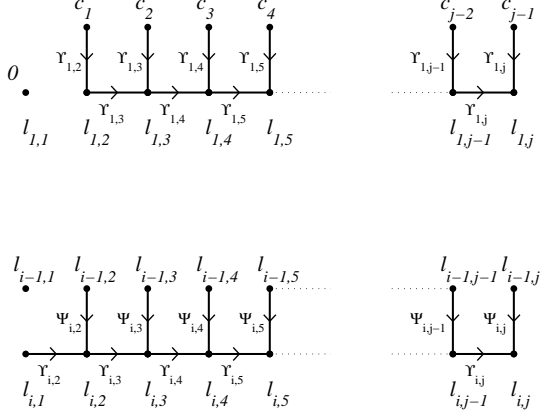


FIG. 10: The components of flux of probability around the link-space matrix. This is general to any attachment kernel. An element within the link-space matrix can be built up from contributing elements and the appropriate factors.

where  $A$  is a constant. This would imply for the degree distribution

$$c_i = \frac{A}{i^2(i+1)}. \quad (\text{B4})$$

Summing over the entire node-space would give  $A = \frac{6}{\pi^2 - 6}$ . However, the boundary conditions (which could be interpreted as influx of probability into the diffusive matrix) are such that this solution does not hold. This is evident when comparing Eq. (B4) with Eq. (B1), which compares to simulated networks well.

We can obtain an exact (although somewhat less pretty) solution by tracing fluxes of probability around the matrix and making use of the previously derived degree distribution. We can rewrite our link-space master equation, (B2), in terms of vertical and horizontal components:

$$\begin{aligned} l_{i,j} &= \Psi_{i,j} l_{i-1,j} + \Upsilon_{i,j} l_{i,j-1}, \\ l_{1,j} &= \Upsilon_{1,j} (l_{1,j-1} + c_{j-1}), \\ l_{1,1} &= 0, \end{aligned} \quad (\text{B5})$$

where

$$\begin{aligned} \Psi_{i,j} &= \frac{\frac{\Theta_{i-1}}{c_{i-1}}}{1 + \frac{\Theta_i}{c_i} + \frac{\Theta_j}{c_j}}, \\ \Upsilon_{i,j} &= \frac{\frac{\Theta_{j-1}}{c_{j-1}}}{1 + \frac{\Theta_i}{c_i} + \frac{\Theta_j}{c_j}}. \end{aligned} \quad (\text{B6})$$

By considering the probability fluxes as shown in Fig. 10, we can write the individual elements in the link-

space matrix as

$$\begin{aligned} l_{1,j} &= \sum_{y=2}^j \left( c_{y-1} \prod_{x=y}^j \Upsilon_{1,x} \right), \\ l_{i,j} &= \sum_{y=2}^j \left( l_{i-1,y} \Psi_{i,y} \prod_{x=y+1}^j \Upsilon_{i,x} \right) + l_{i,1} \prod_{x=2}^j \Upsilon_{i,x}, \\ l_{1,1} &= 0. \end{aligned} \quad (\text{B7})$$

Note that we have yet to introduce the attachment probability kernels and the analysis so far is general. Using the preferential attachment kernel,  $\Theta_i \approx \frac{ic_i}{2}$ , we can write our component-wise factors for the master equation as

$$\begin{aligned} \Psi_{i,j} &= \frac{i-1}{i+j+2}, \\ \Upsilon_{i,j} &= \frac{j-1}{i+j+2}. \end{aligned} \quad (\text{B8})$$

Substituting Eq. (B8) into Eq. (B7) and using the previously derived degree distribution of Eq. (B1) yields for the first row

$$\begin{aligned} l_{1,j} &= \sum_{y=2}^j \left( \frac{4}{y(y-1)(y+1)} \prod_{x=y}^j \frac{x-1}{x+3} \right) \\ &= \frac{4(j-1)!}{(j+3)!} \sum_{y=2}^j (y+2) \\ &= \frac{2(j+6)(j-1)}{j(j+1)(j+2)(j+3)}. \end{aligned} \quad (\text{B9})$$

Subsequent rows can be written as

$$\begin{aligned} l_{i,j} &= l_{i,1} \frac{(j-1)!(3+i)!}{(2+i+j)!} \\ &\quad + \sum_{y=2}^j l_{i-1,y} (i-1) \frac{(j-1)!}{(2+i+j)!} \frac{(1+i+y)!}{(y-1)!} \\ &= \frac{(j-1)!}{(2+i+j)!} \left\{ (3+i)! l_{i,1}, \right. \\ &\quad \left. + (i-1) \sum_{y=2}^j l_{i-1,y} \frac{(1+i+y)!}{(y-1)!} \right\}. \end{aligned} \quad (\text{B10})$$

Rewriting this gives

$$l_{i,j} = \frac{(j-1)!}{(2+i+j)!} \left\{ K_i + E_{i-1,j} \right\}, \quad (\text{B11})$$

where the meaning of  $K_i$  and  $E_{i-1,j}$  follows from Eq. (B10). Clearly, we can write Eq. (B11) for  $l_{i-1,y}$  as

$$l_{i-1,y} = \frac{(y-1)!}{(1+i+y)!} \left\{ K_{i-1} + E_{i-1,y} \right\}. \quad (\text{B12})$$

Substituting Eq. (B12) into Eq. (B10) yields

$$l_{i,j} = \frac{(j-1)!}{(2+i+j)!} \left\{ K_i + (i-1) \sum_{y=2}^j (K_{i-1} + E_{i-1,y}) \right\}. \quad (\text{B13})$$

In order to solve this recurrence relation, we define an operator for repeated summation,  $S_{j,y}^n$ , such that

$$\begin{aligned} S_{j,y}(f(y)) &= \sum_{y=2}^j f(y), \\ S_{j,y}^n(f(y)) &= \sum_{y_n=2}^j \sum_{y_{n-1}=2}^{y_n} \sum_{y_{n-2}=2}^{y_{n-1}} \cdots \sum_{y_3=2}^{y_4} \sum_{y_2=2}^{y_3} \sum_{y=2}^{y_2} f(y). \end{aligned} \quad (\text{B14})$$

The first subscript denotes the initial variable to be summed over and the second the final limit. A few examples of this operation will clarify its use:

$$\begin{aligned} S_{j,y}^0(f(y)) &= f(y), \\ S_{j,y}(1) &= j-1, \\ S_{j,y}^2(1) &= S_{j,y}(y) - S_{j,y}(1) \\ &= \frac{1}{2}(j^2 - j), \\ S_{j,y}^3(1) &= \frac{1}{6}(j^3 - j), \\ S_{j,y}(y) &= \frac{j(j+1)}{2} - 1, \\ S_{j,y}^2(y) &= \frac{1}{6}(j^3 + 3j^2 - 4j). \end{aligned} \quad (\text{B15})$$

We can use this operator in our expression for the element  $l_{i,j}$  in Eq. (B13) and expand to the easily derived value  $E_{2,y}$ :

$$\begin{aligned} l_{i,j} &= \frac{(j-1)!}{(2+i+j)!} \left\{ K_i + (i-1) S_{j,y}(K_{i-1} + E_{i-1,y}) \right\} \\ &= \frac{(j-1)!}{(2+i+j)!} \left\{ K_i + (i-1) K_{i-1} S_{j,y}(1) \right. \\ &\quad \left. + (i-1)(i-2) S_{j,y}^2(K_{i-2} + E_{i-2,y}) \right\} \\ &= \frac{(j-1)!}{(2+i+j)!} \left\{ K_i + (i-1) K_{i-1} S_{j,y}(1) \right. \\ &\quad + (i-1)(i-2) K_{i-2} S_{j,y}^2(1) \\ &\quad + (i-1)(i-2)(i-3) K_{i-3} S_{j,y}^3(1) + \dots \\ &\quad + (i-1)(i-2)(i-3) \times \dots \times 2 K_2 S_{j,y}^{i-2}(1) \\ &\quad \left. + (i-1)(i-2) \times \dots \times 2 S_{j,y}^{i-2}(E_{2,y}) \right\}. \end{aligned} \quad (\text{B16})$$

We can express  $E_{2,y}$  in terms of the operator  $S$  too:

$$E_{2,y} = 4S_{y,y}(2S_{y,y}(1) + S_{y,y}(y)). \quad (\text{B17})$$

The element  $l_{i,j}$  can be expressed as

$$\begin{aligned} l_{i,j} &= \frac{(j-1)!}{(2+i+j)!} \left\{ \sum_{m=2}^i \frac{(i-1)!}{(m-1)!} K_m S_{j,y}^{i-m}(1) \right. \\ &\quad \left. + 4(i-1)!((2S_{j,y}^i(1) + S_{j,y}^i(y))) \right\}. \end{aligned}$$

This form is somewhat obtuse as the calculation of the operator values is less than obvious. However, we can transform to a more easily interpreted operator  $W(n)$  analogous to  $S$  but with different limits such that

$$\begin{aligned} W_{j,y}(f(y)) &= \sum_{y=1}^j f(y), \\ W_{j,y}^n(f(y)) &= \sum_{y_n=1}^j \sum_{y_{n-1}=1}^{y_n} \sum_{y_{n-2}=1}^{y_{n-1}} \cdots \sum_{y_3=1}^{y_4} \sum_{y_2=1}^{y_3} \sum_{y=1}^{y_2} f(y). \end{aligned} \quad (\text{B18})$$

Whilst, at first glance, it looks like little progress has been made, we only need evaluate the repeated operation on initial function  $f(y) = 1$ . This is exactly solvable and we can rewrite in terms of a function  $G(n)$ , dropping the superfluous  $y$  subscripts:

$$\begin{aligned} G_j(n) &= W_j^n(1), \\ G_j(n) &= \begin{cases} \frac{(j+n-1)!}{n!(j-1)!} & \text{for } n \geq 0 \\ 0 & \text{for } n < 0. \end{cases} \end{aligned} \quad (\text{B19})$$

The inductive proof associated with Eq. (B19) can be understood by path counting for some repeating binomial process and is an intrinsic property of the combinatorial choose coefficient. Consider a repeated coin toss over  $x$  steps. The number of ways of achieving  $n+1$  successes after these  $x$  iterations is  $\binom{x}{n+1}$ . Now, the occurrence of this last success could have happened on the  $(n+1)^{\text{th}}$  iteration or the following one, or any of the subsequent iterations till the  $x^{\text{th}}$  one. For the last successful outcome to occur on the  $m^{\text{th}}$  step,  $n$  successful outcomes must have occurred in the previous steps. The number of ways this could have occurred is  $\binom{m-1}{n}$ . Clearly, summing over all possible  $m$  values, the total possible paths resulting in  $n+1$  successes must equate to  $\binom{x}{n+1}$ .

For clarity, this is depicted in Fig. 11. To reach point  $B$  from  $A$  in the binomial process, one of the steps  $w, x, y$  or  $z$  must be traversed, after which there is only one route to  $B$ . Consequently, the number of paths between  $A$  and  $B$  utilising step  $w$  is the same as the number of paths between  $A$  and  $W$ . Similarly, the number of paths between  $A$  and  $B$  utilising step  $x$  is the same as the number of paths between  $A$  and  $X$  and so on. The number of paths between  $A$  and  $B$  can be built expressed as the sum of the paths  $A \rightarrow W, A \rightarrow X, A \rightarrow Y$  and  $A \rightarrow Z$ .

We incorporate this behaviour into our proof for the



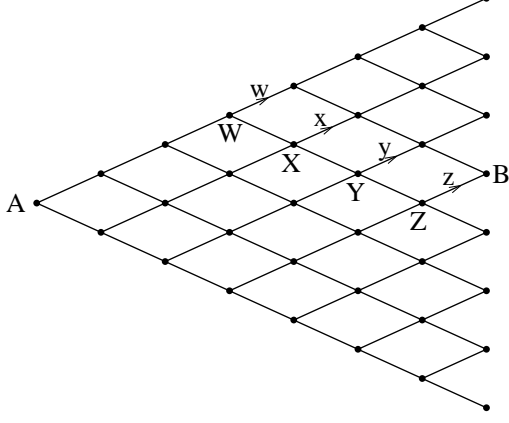


FIG. 11: A binomial process over seven steps. The number of paths between  $A$  and  $B$  can be expressed as the sum of the paths  $A \rightarrow W$ ,  $A \rightarrow X$ ,  $A \rightarrow Y$  and  $A \rightarrow Z$ .

solution of  $G_j(n)$ .

$$\begin{aligned}
 G_j(n) &= \binom{j+n-1}{n}, \\
 G_j(n+1) &= \sum_{i=1}^j G_j(n) \\
 &= \sum_{i=1}^j \binom{j+n-1}{n} \\
 &= \binom{j+n}{n+1}.
 \end{aligned} \tag{B20}$$

As  $G(1) = \binom{j}{1} = j$ , this inductive proof holds for all  $n$  and the following relations hold:

$$\begin{aligned}
 S^n(1) &= G(n) - G(n-1), \\
 S^n(j) &= G(n+1) - G(n-1).
 \end{aligned} \tag{B21}$$

We can now write the element in our link-space matrix for preferential attachment exactly as our function  $G(n)$  is easily evaluated as

$$\begin{aligned}
 l_{i,j} &= \frac{4(j-1)!(i-1)!}{(j+i+2)!} \left( G(i+1) + 2G(i) - 3G(i-1) \right. \\
 &\quad \left. + \frac{1}{2} \sum_{k=2}^i (k-1)(k+6) (G(i-k) - G(i-k-1)) \right).
 \end{aligned} \tag{B22}$$

A comparison of this solution is made to a numerically-populated link-space matrix in Fig. 12.

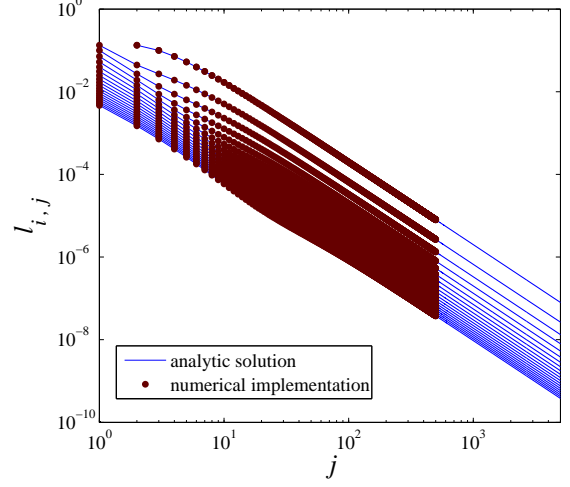


FIG. 12: (Color Online) Comparison of the numerically-derived link-space matrix (markers) and the analytic solution (solid lines) for preferential attachment. The numerical implementation populates the matrix from the solution of the degree distribution and the link-space recurrence equations.

### APPENDIX C: DEGREE DISTRIBUTION OF DECAYING NETWORK

The link-space master equation for the steady state of the decaying networks of Section IV is written as

$$l_{i,j} = \frac{il_{i+1,j} + jl_{i,j+1}}{i+j-2}. \tag{C1}$$

The number of  $1 \leftrightarrow 1$  links does not reach a steady state for this process. However, we can investigate the properties of the rest of the links in the network which do reach a steady state by neglecting these two-node components. In a similar manner to [11] we can make use of a substitution to find a solution to this recurrence equation, namely, for  $i+j \neq 2$  as

$$l_{i,j} = r_{i,j} \frac{(i+j-3)!}{(i-1)!(j-1)!}. \tag{C2}$$

From this, we can obtain

$$r_{i,j} = r_{i+1,j} + r_{i,j+1}. \tag{C3}$$

This has the simple solution:

$$r_{i,j} = \frac{A}{2^{i+j}}. \tag{C4}$$

The link-space for this system can be written for  $i+j \neq 2$  as

$$l_{i,j} = \frac{A}{2^{i+j}} \frac{(i+j-3)!}{(i-1)!(j-1)!}. \tag{C5}$$

Although the summation over this link-space matrix diverges,  $\sum_i \sum_j l_{i,j} \rightarrow \infty$ , we can use the form of Eq. (C5)

to infer the shape of the degree distribution. Recalling that

$$c_i = \frac{M}{N} \frac{\sum_k l_{i,k}}{i}, \quad (\text{C6})$$

we can write for the degree distribution, neglecting the two node components in the network, as

$$\begin{aligned} c_1 &= \frac{MA}{N} \sum_{k=2}^{\infty} \frac{(k-2)!}{(k-1)!2^{1+k}}, \\ c_i &= \frac{MA}{Ni} \sum_{k=1}^{\infty} \frac{(i+k-3)!}{(i-1)!(k-1)!2^{i+k}}, \quad i > 1. \end{aligned} \quad (\text{C7})$$

The  $c_1$  value can be solved by considering the Maclaurin Series of the function  $\log(1+x)$  (as a historical note, this was done by Mercator as early as 1668) and evaluating for  $x = -\frac{1}{2}$ :

$$\begin{aligned} \log(1+x) &= \sum_{k'=1}^{\infty} \frac{(-1)^{k'}}{k'} x^{k'}, \\ \log\left(\frac{1}{2}\right) &= - \sum_{k'=1}^{\infty} \frac{1}{k' 2^{k'}} \\ &= -\log(2). \end{aligned} \quad (\text{C8})$$

Rearranging Eq. (C7) and letting  $k = k' + 1$  yields

$$\begin{aligned} c_1 &= \frac{MA}{N} \sum_{k=2}^{\infty} \frac{1}{(k-1)2^{1+k}} \\ &= \frac{MA}{N} \sum_{k'=1}^{\infty} \frac{1}{k' 2^{k'+2}} \\ &= \frac{MA}{4N} \log(2). \end{aligned} \quad (\text{C9})$$

For node degree greater than one, the solution of (C7) requires a simple proof by induction. Consider the function  $Q(i')$  defined for positive integer  $i'$  as

$$Q(i') = \sum_{k'=0}^{\infty} \frac{\binom{i'+k'}{k'}}{2^{k'+i'}}. \quad (\text{C10})$$

We can write  $Q(i' + 1)$  with similar ease as

$$Q(i' + 1) = \sum_{k'=0}^{\infty} \frac{\binom{i'+k'+1}{k'}}{2^{k'+i'+1}}. \quad (\text{C11})$$

This leads to

$$\begin{aligned} 2Q(i' + 1) - Q(i') &= \sum_{k'=0}^{\infty} \left( \binom{i'+k'+1}{k'} - \binom{i'+k'}{k'} \right) \frac{1}{2^{k'+i'}} \\ &= \sum_{k'=1}^{\infty} \binom{i'+k'}{k'-1} \frac{1}{2^{k'+i'}}. \end{aligned} \quad (\text{C12})$$

A quick substitution of  $k'' = k' - 1$  leads to:

$$\begin{aligned} 2Q(i' + 1) - Q(i') &= \sum_{k''=0}^{\infty} \binom{i'+k''+1}{k''} \frac{1}{2^{k''+i'+1}} \\ &= Q(i' + 1), \\ \Rightarrow Q(i' + 1) &= Q(i'), \\ Q(1) &= 2 \\ &= Q(i'), \end{aligned} \quad (\text{C13})$$

which holds of for all  $i'$ . Rearranging Eq. (C7) and using some simple substitutions,  $i' = i - 2$  and  $k' = k - 1$ , we can derive the following:

$$\begin{aligned} c_i &= \frac{MA}{Ni} \sum_{k=1}^{\infty} \frac{(i+k-3)!}{(i-1)!(k-1)!2^{i+k}} \\ &= \frac{MA}{Ni(i-1)} \sum_{k=1}^{\infty} \frac{(i+k-3)!}{(i-2)!(k-1)!2^{i+k}} \\ &= \frac{MA}{Ni(i-1)} \sum_{k'=0}^{\infty} \frac{\binom{i'+k'}{k'}}{2^{i'+k'+3}} \\ &= \frac{MA}{4N} \frac{1}{i(i-1)}. \end{aligned} \quad (\text{C14})$$

We can normalize this degree distribution such that  $\sum_i c_i = 1$  and so for the decaying network without the two node components (the  $1 \leftrightarrow 1$  links), the degree distribution can be expressed as

$$\begin{aligned} A &= \frac{4N}{M(1 + \log(2))}, \\ c_1 &= \frac{\log(2)}{1 + \log(2)}, \\ c_i &= \frac{1}{(1 + \log(2))i(i-1)}, \quad i > 1. \end{aligned} \quad (\text{C15})$$

- 
- [1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47–97 (2002).  
[2] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).  
[3] M.E.J. Newman, SIAM Rev. **45**, 167–256 (2003).  
[4] S. Boccaletti, V. Latorab, Y. Morenod, M. Chavez and

- D.-U. Hwanga, Physics Reports **424**, 175–308 (2006).  
[5] M.E.J. Newman, A.L. Barabási and D.J. Watts, *The Structure and Dynamics of Networks* (Princeton University Press, Princeton, 2006).  
[6] L. da F. Costa, F. A. Rodrigues, G. Travieso and P. R. Villas Boas, Advances in Physics, **56**, 167–242 (2007).

- [7] A. Barrat and R. Pastor-Satorras, Phys. Rev. E **71**, 036127 (2005).
- [8] A.L. Barabási and R. Albert, Science **286**, 509–512 (1999).
- [9] A.L. Barabási, R. Albert and H. Jeong, Physica A **272**, 173–187 (1999).
- [10] E. Erdős, and A. Rényi, *Publ. Math. Debrecen* **6**, 290–297 (1959).
- [11] P.L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
- [12] S.N. Dorogovtsev, J.F.F. Mendes and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633–4636 (2000).
- [13] K. Klemm and V.M. Eguíluz, Phys. Rev. E **65**, 057102 (2002).
- [14] A. Fronczak, P. Fronczak and J.A. Holyst, Phys. Rev. E **68**, 046126 (2003).
- [15] D.J. Watts and S.H. Strogatz, Nature **393**, 440 (1998).
- [16] P. Holme and B.J. Kim, Phys. Rev. E **65**, 026107 (2002).
- [17] R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen and K. Kaski, Physica A **371**, 851–860 (2006).
- [18] G. Szabó, M. Alava and J. Kertész, Phys. Rev. E **67**, 056102 (2003).
- [19] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **68**, 036112 (2003).
- [20] R. Pastor-Satorras, A. Vázquez and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).
- [21] M.E.J. Newman Phys. Rev. Lett. **89**, 208701 (2002).
- [22] S.N. Soffer and A. Vázquez, Phys. Rev. E **71**, 057101 (2005).
- [23] S. Maslov, K. Sneppen and A. Zaliznyak, cond-mat/0205379, (2002).
- [24] J. Park and M.E.J. Newman, Phys. Rev. E **68**, 026112 (2003).
- [25] M. Catanzaro, G. Caldarelli and L. Pietronero, Phys. Rev. E **70**, 037101 (2004).
- [26] V. M. Eguíluz and K. Klemm, Phys. Rev. Lett. **89**, 108701 (2002).
- [27] M. Boguñá, R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **90** 028701 (2003).
- [28] M. Boguñá and R. Pastor-Satorras, Phys. Rev. E **66**, 047104 (2002).
- [29] M. Brede and S. Sinha, cond-mat/0507710 (2005).
- [30] M. di Bernardo, F. Garofalo and F. Sorrentino, cond-mat/0504335 (2006).
- [31] M. di Bernardo, F. Garofalo and F. Sorrentino, cond-mat/0506236 (2005).
- [32] S. Maslov and K. Sneppen, Science **296**, 910 (2002).
- [33] J. Berg, M. Lässig and A. Wagner, BMC Evolutionary Biology **4**, 51 (2004).
- [34] P.L. Krapivsky, S. Redner and F. Leyvraz, Phys. Rev. Lett. **85**, 4629–4632 (2000).
- [35] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman and S. H. Strogatz, Phys. Rev. E **64**, 041902 (2001).
- [36] D.M.D Smith, J.-P. Onnela and N.F. Johnson, New Journal of Physics, **9** 191 (2007).
- [37] Bollobás, B., Discrete Mathematics **33** 1–19 (1981).
- [38] For mathematical details, derivations and further information, see D.M.D. Smith, C.-F. Lee, N.F. Johnson and J.-P. Onnela, physics/0702010 (2006).
- [39] A. Vázquez, R. Pastor-Satorras and A. Vespignani, cond-mat/00112400 (2001).
- [40] N. Farid and K. Christensen, New Journal of Physics **8** 212 (2006).
- [41] S. Laird and H.J. Jensen, *Europhys. Lett.* **76**, 710–716 (2006).
- [42] M. Salathé, R.M. May and S. Bonhoeffer, R.J. R. Soc. Interface, **2**, 533–536 (2005).
- [43] A. Vázquez, Phys. Rev. E **67**, 056104 (2003).
- [44] T.S. Evans and J.P. Saramäki, Phys. Rev. E **72**, 026138 (2005).
- [45] J.P. Saramäki and K. Kaski, Physica A **341**, 80–86 (2004).
- [46] V. Batagelj and A. Mrvar, Connections **21**, 47–57 (1998).
- [47] The master equation does not represent the evolution of an ensemble average of networks. Each specific realisation will have its own evolution of attachment kernel which cannot be described by the ensemble average. The existence of a steady state to the master equation does not necessarily imply that specific realisations converge upon it in the large  $N$  limit but simply that the solution is static with respect to the growth algorithm. In practice, however, a master equation approach often yields good analytic agreement with even single realisations [38].
- [48] A network can be constructed for an arbitrary, normalized link-space matrix. A diagonal element  $l_{i,i}$  can be simply obtained through the addition of fully connected components of  $i + 1$  nodes. Similarly an off-diagonal element  $l_{i,j}$  can be generated through suitable rewiring of  $i$ -degree and  $j$ -degree cliques.
- [49] When new links are allowed between existing nodes, the exponent can be as low as 2 for a mixture of random and preferential attachment [44].
- [50] A random walk of two steps in a network will be biased towards selecting low degree nodes. Previous studies have made use of this phenomenon when employing a two (or more) step random walk to select two (or more) nodes within the existing network for a new node to attach to with two (or more) links. The opposing biases effectively cancel such that the resulting algorithm can replicate preferential attachment. This was studied by Saramäki and Kaski [45] and by Evans and Saramäki [44]. For a more detailed analysis see [38].